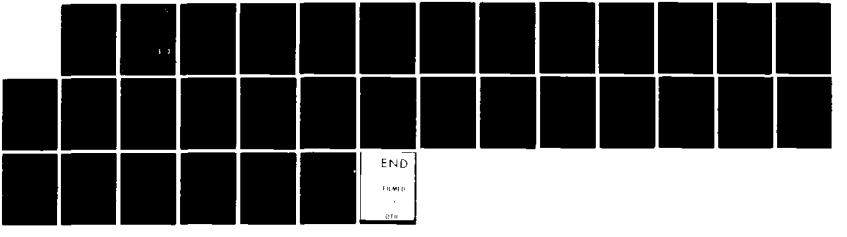
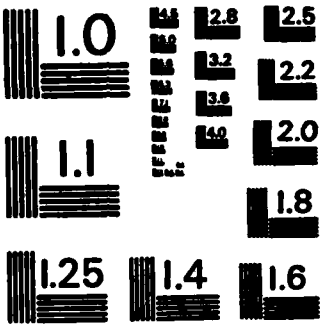
AD-A121 874 FACTOR SCREENING IN SIMULATION: EVALUATION OF TWO STRATEGIES ON BASED ON.. (U) DESMATICS INC STATE COLLEGE PA C A MAURO ET AL. NOV 82 TR-113-7 N00014-79-C-0650 1/1

UNCLASSIFIED F/G 12/1 NL





FACTOR SCREENING IN SIMULATION: EVALUATION OF TWO STRATEGIES BASED ON RANDOM BALANCE SAMPLING

by

Carl A. Mauro
and
Dennis E. Smith

— STATISTICS —

— OPERATIONS RESEARCH —

— MATHEMATICS —

DESMATICS, INC.

P.O. Box 618
State College, PA 16801

82 11 29 034

DESMATICS, INC.

P. O. Box 618
State College, Pa. 16801
Phone: (814) 238-9621

Applied Research in Statistics - Mathematics - Operations Research

FACTOR SCREENING IN SIMULATION: EVALUATION OF TWO STRATEGIES BASED ON RANDOM BALANCE SAMPLING

by

Carl A. Mauro
and
Dennis E. Smith

TECHNICAL REPORT NO. 113-7

November 1982

This study was supported by the Office of Naval Research under Contract No. N00014-79-C-0650, Task No. NR 042-467

TABLE OF CONTENTS

1. INTRODUCTION

Computer simulation, an extremely useful and popular technique in operations research and management science, is often used to study the behavior of very complex real-world systems. Unfortunately, simulation models of complex systems tend to be extremely complicated themselves because of the meticulous detail ordinarily included in such models. Furthermore, computer codes corresponding to these models are usually extraordinarily large and very long-running.

Often, simulation users cannot readily assimilate the information contained in large, complex codes because they are overwhelmed by the vast number of factors (i.e., input variables) and are confused about how to make an effective analysis of the model without having to perform an excessive number of costly and time-consuming simulation runs. If the users could identify the most important factors in some reasonable way, they could make the model more manageable and their analysis more efficient by concentrating the major experimental effort on the key factors.

Factor screening methods are statistical methods that attempt to identify the more important variables. (See, for example, [4], [6], and [8].) A basic function of these methods is to sort all the factors into two primary groups. One group consists of the "important" factors which are judged worthwhile to investigate further, while the other consists of the remaining "unimportant" factors.

When selecting an appropriate factor screening method, one must primarily consider the number of runs available for screening. In the simulation environment, the number of factors to be screened almost always ex-

ceeds the available number of simulation runs. In statistical experimental design terminology, this is known as a supersaturated situation. Such a situation is common in the simulation framework because of the large number of factors usually under consideration and because of the time and cost of the computer runs.

Although many strategies have been suggested for designing and conducting screening experiments, few are applicable to the supersaturated case. Furthermore, for those few that are, there has been no systematic evaluation and comparison of their performance. In this paper we provide quantitative information on a supersaturated screening strategy based on random balance sampling ([1], [10], [12]). In addition, we compare this strategy with a modified strategy based on a combination of random balance and Plackett-Burman designs [9].

2. PRELIMINARY DISCUSSION

To provide a common statistical basis to compare and assess screening strategies, we must make some assumptions as to the general structure of a simulation model. For detecting the factors having major effects it is usually reasonable to assume

$$y_i = \beta_0 + \sum_{j=1}^{K} \beta_j x_{ij} + \varepsilon_i \quad , \tag{2.1}$$

where y_i is the value of the response (i.e., output variable) in the $i^{\underline{th}}$ simulation run; K is the total number of factors to be screened, each of which is at two levels (±1); $x_{ij} = \pm 1$ depending on the level of the $j^{\underline{th}}$ factor during the $i^{\underline{th}}$ simulation run; β_j is the (linear) effect of the $j^{\underline{th}}$ factor; and the error terms ε_i are independent and normally distributed random disturbances with zero mean and variance σ^2.

In essence, model (2.1) is a first-order Taylor series approximation to an actual relationship between output and input variables; ordinarily we would use this approximation over a relatively small region of the factor space. We will restrict performance evaluation to this model.

Random Balance Sampling

An experiment involving random balance sampling is based on an experimental design that is random. In a two-level (±1) random balance design, each column of the design matrix consists of N/2 +1's and N/2 -1's where N (an even number) denotes the total number of runs to be made. The +1's and -1's in each column are assigned randomly, making all possible combinations of N/2 +1's and N/2 -1's (there are $C^N_{N/2}$ in all) equally likely, with each

column receiving an independent randomization.

The principal advantage of random balance (RB) sampling for use in screening is its flexibility. We can select N independently of K; there is no mathematical restriction or relationship between N and K as there is in more traditional experimental designs. A second advantage is the ease with which we can prepare RB designs for any N and K, an important consideration when K is large.

There are two corresponding disadvantages to RB sampling. The first of these is that factors are confounded to a random degree. Thus, we cannot generally control the amount of confounding or interdependence between factors. Secondly, there is no specialized or unique technique for analyzing RB designs. The simplest approach is to consider each factor separately and apply a standard F-test. We should mention, however, that practically any technique used to analyze data without RB properties can be used to analyze any (sufficiently small) subset of factors in an RB design. This is done by simply ignoring any factor not included in the particular set of factors being analyzed.

In this paper, we consider a standard F-test applied separately to each factor as the method of analysis for random balance data. Furthermore, for simplicity, we conduct each F-test at the same level of significance α_1. An RB strategy, therefore, is completely determined once we specify N and α_1. Accordingly, we denote such a strategy by RB(N,α_1). Moreover, we classify a factor as important only if it has a significant F-ratio.

Modified Strategy

We now consider a two-stage screening strategy having an RB(N,α_1)

first-stage followed by the use of a second-stage Plackett-Burman (PB) design. We include a given factor in the second-stage PB experiment only if it has a significant F-ratio in the first-stage RB experiment. In this combination strategy, we declare important those factors which reach and have a significant effect in the second stage.

Because PB designs are orthogonal, the second stage separates any confounding between factors carried over from the RB first stage. Factors not formally included in the second-stage experiment are held at a constant level so not to bias any of the second-stage estimates. Further, unlike RB designs, we can analyze PB designs by the usual analysis of variance procedures for factorial experiments. We denote this combination strategy by RP(N,α_1,α_2) where α_2 is the significance level used in all second-stage F-tests.

The total number of runs R required by an RP strategy will therefore be N+M where M denotes the number of second-stage runs. Although we can specify N, the number of second-stage runs M will depend on the number of factors S carried over from the first stage. For reasons of economy and to avoid design saturation (i.e., no degrees of freedom to estimate experimental error), we employ the smallest PB design that guarantees at least one error degree of freedom. Since PB designs are only available for numbers of runs that are multiples of four, we can obtain a minimum of one and a maximum of four error degrees of freedom by following this convention.

Accordingly, we can write M mathematically as M = B(S+1) where B(x) = x+4-x(mod 4). Thus, R = N+M = N+B(S+1). We should emphasize, however, that R is random (since M is). Hence, in an RP(N,α_1,α_2) strategy we do not know prior to experimentation the exact number of runs that will be required. This, of course, is a disadvantage of the RP strategy. However, noting that

$B(x) \approx x + 2.5$ we can approximate $E(R)$ by

$$E(R) \approx N + E(S) + 3.5 \ . \tag{2.2}$$

Since $|B(x) - (x+2.5)| \leq 1.5$, the approximation in (2.2) can differ from $E(R)$ by at most 1.5 runs. In Section 3 we will show how the quantities N, α_1, and α_2 affect the performance of the RB and RP screening strategies.

3. PERFORMANCE ASSESSMENT

In general, the objectives of a factor screening strategy are (1) to detect as many important factors as possible, (2) to declare important as few unimportant factors as possible, and (3) to perform as few runs as possible. In short, one must consider both how many runs a strategy requires and how accurately it classifies factors. It is difficult, however, to dichotomize factors as either important or unimportant. From a practical standpoint, the importance (or unimportance) of a factor will depend on the magnitude of its effect relative to that of experimental error, σ, and that of the magnitudes of other effects present. Importance, therefore, is essentially a matter of degree. The greater (lesser) the degree of importance, the larger should be the probability of classifying the factor as important (unimportant).

In this section we provide formulas that summarize the performance of the $RB(N,\alpha_1)$ and $RP(N,\alpha_1,\alpha_2)$ strategies in terms of the number (or expected number) of runs a strategy requires and in terms of a strategy's sensitivity (i.e., power) for declaring a factor important. In order to compare strategies, we should note that tradeoffs will need to be made. Indeed, objectives (1) and (2), which deal with factor classification, conflict with objective (3), which deals with testing cost. In many ways the screening problem is like the testing of a statistical hypothesis in which we want the sample size to be small but the power (i.e., the probability of rejecting a false null hypothesis) to be large. Our intent is to provide the simulation user with quantitative information on the tradeoffs involved.

To establish some notation, we define $R_{1j}=1$ if the $j^{\underline{th}}$ factor is declared important by an $RB(N,\alpha_1)$ strategy, and we define $R_{2j}=1$ if the $j^{\underline{th}}$

factor is declared important by an RP(N,α_1,α_2) strategy; otherwise, we let R_{1j}=0 and R_{2j}=0. Except in the simplest cases, P(R_{1j}=1) and P(R_{2j}=1) are too complex to be evaluated analytically. In lieu of exact solutions, we develop approximations to these probabilities. In addition, we present an approximation to the expected value of R, the total number of runs required by an RP(N,α_1,α_2) strategy. The total number of runs required by an RB(N,α_1) strategy is, of course, N runs.

The RB Strategy

We can write model (2.1) in matrix terms as $y=\beta_0\underline{1} + \underline{X}\beta + \underline{\varepsilon}$ where $\underline{1}$ is an N x 1 vector of +1's, $\underline{y}$ is an N x 1 vector of responses, $\underline{\varepsilon}$ is an N x 1 vector of error terms, $\underline{\beta}$ is a K x 1 vector of factor effects, and $\underline{X}$ is an N x K design matrix. In a random balance experiment $\underline{X}=[\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_K]$ is a stochastic matrix whose j^{th} column, $\underline{x}_j$, is an N x 1 vector consisting of a random arrangement of N/2 +1's and N/2 -1's. By construction, the K column vectors of $\underline{X}$ are independent. We assume that $\underline{X}$ and $\underline{\varepsilon}$ are independent.

The simple least squares estimator of β_j ($j\geq 1$) is given by

$$\hat{\beta}_j = (\bar{y}_{+j}-\bar{y}_{-j})/2 \qquad (3.1)$$

where $\bar{y}_{+j}$ ($\bar{y}_{-j}$) is the average value of the response over the N/2 runs at the +1 (-1) level of the j^{th} factor. (By simple least squares we mean that each β_j is estimated igoring all other factors.) In matrix terms

$$\hat{\beta}_j = (\underline{x}_j'\underline{y})/N = [(\sum_{i=1}^{K} \beta_i \underline{x}_j'\underline{x}_i) + \underline{x}_j'\underline{\varepsilon}]/N \; . \qquad (3.2)$$

Thus,
$$E(\hat{\beta}_j) = (1/N)[\sum_{i=1}^{K} \beta_i E(\underline{x}_j'\underline{x}_i) + E(\underline{x}_j'\underline{\varepsilon})], \qquad (3.3)$$

and $$V(\hat{\beta}_j) = (1/N^2)\left[\sum_{i=1}^{K}\beta_i{}^2 V(\underline{x}_j'\underline{x}_i) + V(\underline{x}_j'\underline{\varepsilon})\right] \tag{3.4}$$

where in (3.4) we make use of the fact that $\underline{x}_j'\underline{\varepsilon}, \underline{x}_j'\underline{x}_1, \underline{x}_j'\underline{x}_2, \ldots, \underline{x}_j'\underline{x}_K$ are mutually independent for fixed j.

It is clear that the exact sampling distribution of $\hat{\beta}_j$ is intractable. However, using results we derive in the Appendix, we can easily show

$$E(\hat{\beta}_j) = \beta_j, \tag{3.5}$$

$$V(\hat{\beta}_j) = (\tau^2 - \beta_j^2)/(N-1) + \sigma^2/N, \tag{3.6}$$

and $$\text{cov}(\hat{\beta}_i, \hat{\beta}_j) = \beta_i\beta_j/(N-1) \tag{3.7}$$

where $\tau^2 = \sum_{m=1}^{K}\beta_m^2$. The simple least squares estimator defined in (3.1) is therefore an unbiased estimator of β_j, although its variance can be seriously inflated. Moreover, the correlation between $\hat{\beta}_i$ and $\hat{\beta}_j$ is roughly

$$\text{corr}(\hat{\beta}_i, \hat{\beta}_j) \approx \beta_i\beta_j/\phi_i\phi_j \tag{3.8}$$

where $\phi_m^2 = \tau^2 + \sigma^2 - \beta_m^2$.

The correlation in (3.8) is a measure of the confounding between $\hat{\beta}_i$ and $\hat{\beta}_j$. We make the somewhat surprising observation that an increase in N does not decrease the confounding in an RB design where simple least squares is used as the estimation method. Furthermore, the degree of confounding between $\hat{\beta}_i$ and $\hat{\beta}_j$ is dependent upon σ and the magnitudes of the other effects in the model.

Regarding formal significance testing, we note that a single-factor F-test to test whether $\hat{\beta}_j$ is significantly different from zero is equivalent to a simple two-sample t-test between the high (+1) and the low (-1) levels of the j^{th} factor. The associated test statistic t_j is given by

$$t_j = \hat{\beta}_j / [SSE_j / N(N-2)]^{\frac{1}{2}} \tag{3.9}$$

where SSE_j is the familiar analysis of variance notation for the error sum of squares of factor j. Computationally,

$$SSE_j = \sum_H (y_i - \bar{y}_{+j})^2 + \sum_L (y_i - \bar{y}_{-j})^2 \tag{3.10}$$

where the first (second) summation is taken over the N/2 observations at the high (low) level of the $j^{\underline{th}}$ factor. An alternative, more direct computational formula is given by

$$SSE_j = \sum_{i=1}^{N} y_i^2 - N\hat{\beta}_j^2 - N\bar{y}^2 , \tag{3.11}$$

where $\bar{y}$ is the overall mean of the N responses.

Following normal theory we reject $H_0: \beta_j = 0$ in favor of $H_1: \beta_j \neq 0$ if the observed value of $|t_j|$ equals or exceeds $t(N-2; \alpha_1/2)$, the upper $100(1-\alpha_1/2)$ percentage point of "Student's" t-distribution having (N - 2) degrees of freedom. Assuming that normal theory is adequate, we can approximate the distribution of t_j with a noncentral t-distribution having (N - 2) degrees of freedom and noncentrality parameter

$$\delta_{1j} = N^{\frac{1}{2}} \beta_j / \phi_j . \tag{3.12}$$

Therefore,

$$P(R_{1j} = 1) \approx \psi_1(\delta_{1j}) , \tag{3.13}$$

where $$\psi_1(\delta) = P\{|T_{N-2}(\delta)| \geq t(N-2;\alpha_1/2)\}$$

and $T_\gamma(\delta)$ denotes a random variable having a noncentral t-distribution with γ degrees of freedom and noncentrality parameter δ.

The RP Strategy

To evaluate $P(R_{2j}=1)$ in an $RP(N,\alpha_1,\alpha_2)$ strategy, we observe first that

$$P(R_{2j}=1) = P(R_{1j}=1)\ P(R_{2j}=1 \mid R_{1j}=1) \tag{3.14}$$

and note that $S = \sum_{j=1}^{K} R_{1j}$, where as defined previously S denotes the number of factors carried over from the $RB(N,\alpha_1)$ first stage to the PB second stage.

Next, we define $S_j = S - R_{1j}$ and write

$$\begin{aligned} P(R_{2j}=1|R_{1j}=1) &= \sum_{s=1}^{K} P(R_{2j}=1,\ S_j=s-1|R_{1j}=1) \\ &= \sum_{s=1}^{K} P(S_j=s-1|R_{1j}=1)\ P(R_{2j}=1|S_j=s-1,\ R_{1j}=1) \\ &= \sum_{s=1}^{K} P(S_j=s-1|R_{1j}=1) P(R_{2j}=1|S=s,\ R_{1j}=1)\ . \end{aligned} \tag{3.15}$$

Using well-known testing properties of PB designs,

$$P(R_{2j}=1|S=s,\ R_{1j}=1) = \psi_2(s,\delta_{2j})\ , \tag{3.16}$$

where $$\psi_2(s,\delta) = P\{|T_{d(s)}(\delta)| \geq t(d(s);\ \alpha_2/2)\}\ , \tag{3.17}$$

$$\delta_{2j} = [B(s+1)]^{\frac{1}{2}}\ \beta_j/\sigma\ , \tag{3.18}$$

and $$d(s) = B(s+1) - (s+1)\ . \tag{3.19}$$

The conditional probability in (3.16) represents the probability that the $j^{\underline{th}}$ factor tests significant from zero in the PB second stage given that it and $s-1$ other factors test significant from zero in the $RB(N,\alpha_1)$ first stage. The quantity d(s) represents the number of error degrees of freedom in a PB design for the study of s factors in $B(s+1)$ runs.

Summarizing up to this point, we have

$$P(R_{2j}=1)=\psi_1(\delta_{1j})\sum_{s=1}^{K}\psi_2(s,\delta_{2j})\,P(S_j=s-1|R_{1j}=1)\ . \tag{3.20}$$

To complete our evaluation of $P(R_{2j}=1)$, we must somehow approximate the conditional distribution of S_j given $R_{1j}=1$ and substitute into equation (3.20). This, however, is a difficult task; the conditional distribution of S_j given $R_{1j}=1$ is extremely complex. Nevertheless, the conditional distribution of S_j given $R_{1j}=1$ might be reasonably approximated, for moderately sized N, as the convolution of $(K-1)$ independent Bernoulli random variables having success probabilities $\{\psi_1(\delta_{1m}):m=1,2,\ldots,j-1,j+1,\ldots,K\}$. Alternatively, following Feller [3], for large N and moderate values of

$$\lambda_j=\sum_{m\neq j}^{K}\psi_1(\delta_{1m})\ ,$$

we might reasonably approximate the conditional distribution of S_j given $R_{1j}=1$ with a Poisson distribution having mean λ_j. The Poisson approximation approach is generally much easier to apply than the Bernoulli convolution approach, particularly if K is large.

Regarding the expected number of runs required by an $RP(N,\alpha_1,\alpha_2)$ strategy, we observe that

$$E(S)=\sum_{j=1}^{K}E(R_{1j})\approx\sum_{j=1}^{K}\psi_1(\delta_{1j})\ . \tag{3.21}$$

Whereupon, introducing (3.21) into (2.2), we obtain

$$E(R)\approx N+3.5+\sum_{j=1}^{K}\psi_1(\delta_{1j})\ . \tag{3.22}$$

Monte Carlo Results

As a check on the various approximations presented in this section, we conducted two Monte Carlo case studies, the results of which are summarized

in Tables 1 and 2. As can be seen from these tables, the results are extremely encouraging and suggest that the approximations of $P(R_{1j} = 1)$, $P(R_{2j} = 1)$, and $E(R)$ given in (3.13), (3.20), and (3.22), respectively, are quite reasonable for practical purposes. It is the authors' experience [5] that even for relatively small values of N these approximations are fairly reasonable.

As can be noted from Table 1, the approximations to $P(R_{2j} = 1)$ based on the Bernoulli convolution and Poisson distribution approaches yield essentially the same results. Because of this agreement and the complexity of the calculations associated with the Bernoulli convolution approach for the second case study, we used only the Poisson distribution approach in the second example to evaluate $P(R_{2j} = 1)$. In this case, nonzero factor effects vary in magnitude, ranging between σ and 6σ in absolute magnitude. Moreover, as might be expected in an actual screening situation, the majority of the effects are relatively small.

4. PRACTICAL IMPLICATIONS AND CONSIDERATIONS

In this section we discuss a number of practical considerations and implications regarding the RB and RP screening strategies. In addition, using a hypothetical situation, we illustrate numerically a direct application of the results of Section 3.

We note, first, that an increase in N or α_1 increases the power of an RB(N,α_1) strategy. Of course, an increase in N also increases experimental costs by requiring more screening runs, and an increase in α_1 increases correspondingly the probability of declaring important a negligible factor.

We should further point out that applying an F-test separately to each factor is not necessarily the most powerful method of analyzing data from a random balance experiment. Presumably, more sophisticated statistical techniques (such as least squares stepwise or stagewise methods) that analyze more than one factor at a time would provide greater power. However, in some applications such methods may not be computationally feasible, particularly if K is exceedingly large, and can be severely limited if N is very small relative to K. The individual F-test approach is generally a relatively quick and easy testing procedure. Moreover, the use of this approach admits to a tractable quantitative assessment, whereas more sophisticated analytical techniques lead to an intractable problem. In any event, we suspect that the results of Section 3 can be regarded as a lower bound for the discriminatory power of alternative, more sophisticated RB analysis and testing procedures.

To increase the power of an RP(N,α_1,α_2) strategy, adjustments can be made in the RB(N,α_1) first stage, the PB second stage, or both. Although an increase in α_1 does not affect the number of runs in an RB(N,α_1) strategy,

this is not true for an RP(N,α_1,α_2) strategy. In this case, an increase in α_1 will usually increase the number of second-stage runs and hence the total number of runs.

To increase the power of a PB experiment, one can employ a larger PB design or a larger level of significance α_2 in the analysis. The power of the second-stage analysis might also be increased if some factor effects could be reasonably assumed to be negligible based on an examination of the data. In such cases, one could pool the sum of squares associated with these factors into the error sum of squares, thus obtaining a pooled error estimate having more degrees of freedom for error than the unpooled estimate. If the pooled effects are indeed negligible, the increased error degrees of freedom will translate into greater power. Pooling is especially appealing when there is only one degree of freedom for error. Caution, however, should be exercised since pooling tends to diminish the denominator expected mean square of the F-ratio.

To help determine which effects, if any, one might reasonably combine into a pooled error estimate, estimated effects can be plotted on normal probability paper. In this technique (e.g., Daniel [2]), negligible effects should fall approximately along a straight line, while large effects should tend to fall far from the line. Because the interpretation of the results relies heavily on subjective judgement, a quantitative assessment of the merits of normal plotting is not possible, however.

In Section 2 we noted one advantage of an RB strategy compared with an RP strategy. This was that for an RB strategy, unlike for an RP strategy, the number of required screening runs is fixed rather than random. Other than making this simple observation it is difficult to make any further general statements or recommendations regarding the usage of these two

strategies relative to each other without further study. Toward this end, we have conducted a study of this problem under some simplified assumptions.

We wish to consider the case in which of the K factors to be screened k have the same absolute effect, say $\Delta > 0$, and the remaining K-k are inactive, that is, have a zero effect. In this case, we can identify three basic measures of performance: c, α, and ψ, where we define

$$c = (\text{total expected number of screening runs})/K \ , \tag{4.1}$$

$$\alpha = P\{\text{declare important an inactive factor}\} \ , \tag{4.2}$$

and $$\psi = P\{\text{declare important an active factor}\} \ . \tag{4.3}$$

By active, we mean one of the k factors having an absolute effect equal Δ. For future reference, we define $p = k/K$.

It is clear that for an $RB(N,\alpha_1)$ strategy $c = N/K, \alpha = \alpha_1$, and ψ can be determined from (3.13). For an $RP(N,\alpha_1,\alpha_2)$ strategy $c = E(R)/K, \alpha = \alpha_1\alpha_2$, and ψ and E(R) can be determined from (3.20) and (3.22), respectively. We note that specifying c and α determines uniquely the corresponding RB strategy, namely, $RB(cK,\alpha)$ where cK is assumed to be an even integer. With this in mind, we developed a computer search routine which for any given $RB(ck,\alpha)$ strategy attempts to find an RP strategy having the same c and α, but greater power.

Using our search routine we examined the following twelve cases: $K = 100, 200, 500$; $p = .05, .15$; and $\Delta/\sigma = 2, 8$. Furthermore, to maintain a supersaturated situation, we considered $.1 \leq c \leq .9$.

Figure 1 presents graphically the results of the investigation. Each curve in this figure represents one of the six combinations of K and p studied. For points (c,α) above each curve, no $RP(N,\alpha_1,\alpha_2)$ strategy could improve on the performance of the corresponding $RB(cK,\alpha)$ strategy. For points (c,α) below each curve, there exists some $RP(N,\alpha_1,\alpha_2)$ strategy that

outperforms the corresponding RB(cK,α) strategy. Each curve defines the boundaries of what we refer to as "zones of inadmissibility." As can be seen from Figure 1, the α level defining each boundary curve increases as sample size increases.

Furthermore, for the cases examined, we found that $r = \Delta/\sigma$ had virtually no effect on the results. We can offer two reasons for this. First, in an RB experiment where all nonzero effects are of equal absolute magnitude, the noncentrality parameter, δ_1, associated with these factors is given by $\delta_1^2 = N/(k - 1 + r^{-2})$, per (3.12). If r is large relative to $(k-1)^{-\frac{1}{2}}$, then $\delta_1^2 \approx N/(k-1)$. Second, in a PB experiment the noncentrality parameter associated with factors of equal absolute effects is given by $\delta_2 = M^{\frac{1}{2}}r$ where M is the number of runs made, see (3.18). For the cases we considered, M is generally large enough that whether r = 2 or r = 8 makes little difference in any resulting power calculation.

In sum, the results indicate that neither strategy dominates the other. Moreover, the findings suggest that an RP strategy should be considered in those situations where it is important that α be maintained at a low level.

Finally, to illustrate a direct application of the results of Section 3, consider a simulation model consisting of K = 200 factors and suppose we are interested in employing an RB(N,α_1) screening strategy. The parameters N and α_1 are at our disposal, of course. Suppose further that we anticipate the average absolute effect to be roughly 0.5σ with a standard deviation of 1.5σ. We imagine that the absolute magnitudes of the effects have a relative frequency distribution similar to that illustrated in Figure 2.

It is convenient to define $\gamma_j = |\beta_j|/\sigma$, the ratio of the j^{th} absolute effect to σ. In terms of the γ_j we can write the square of the noncentrality parameter associated with the j^{th} effect, see (3.12), as

$$\delta_{1j}^2 = N\gamma_j^2/(\sum_{i=1}^{K} \gamma_i^2 + 1 - \gamma_j^2) \quad . \tag{4.4}$$

From our above assumptions, $\Sigma\gamma_i/200 = .5$ and $\Sigma(\gamma_i - .5)^2/200 = (1.5)^2$, from which it follows that $\Sigma\gamma_i^2 = 200(2.25 + .5^2) = 500$. Therefore,

$$\delta_{1j}^2 = N\gamma_j^2/(501 - \gamma_j^2) \quad . \tag{4.5}$$

Of course the larger is δ_{1j}^2, the greater is the power of the testing procedure, see (3.13). Further, it is clear that if $|\beta_i| \leq |\beta_j|$, then $\delta_{1i}^2 \leq \delta_{1j}^2$.

To make any power calculations, we need to specify a value or values of γ_j. For the sake of illustration, suppose we wish to consider the power for an effect of magnitude 3σ, that is, consider a $\gamma_j = 3$. Accordingly, introducing $\gamma_j = 3$ into (4.5), we have $\delta_{1j}^2 = 9N/492$. Therefore, for a given N and α_1, we can compute the power.

Table 3 presents the power corresponding to an effect of magnitude 3σ for a number of RB(N,α_1) strategies. (From a previous remark, we know that power will be greater for any effect greater than 3σ in absolute magnitude.) We can use Table 3 for guidance in using and selecting a suitable RB(N,α_1) strategy. For example, if we decide to control α_1 at .15 and desire our power to be at least 0.50 for effects greater than or equal 3σ in absolute magnitude, then N, by interpolation, must be at least 118 runs. If an N this size is not feasible, then additional tradeoffs must be made or we should investigate the use of an alternative screening plan.

The use of an RP(N,α_1,α_2) strategy could be investigated in a similar manner, again utilizing the results of Section 3. Here, of course, the effect of three parameters, not two, would have to be considered.

5. SUMMARY AND DISCUSSION

In this paper we have focused on the problem of factor screening (i.e., the identification of the important variables) in computer simulation. Specifically, we have discussed and evaluated two factor screening strategies, both of which are based on random balance sampling. These two strategies are intended for use in supersaturated situations, that is, where the number of variables to be screened exceeds the number of runs available. This type of situation frequently exists in the simulation environment, particularly in the study of large, complex models.

Generally, in any screening application we must consider both the accuracy of factor identification and the number of runs required by the strategy we employ. For the two strategies we have considered, we have developed approximations to (1) the probability that a given factor is declared important and (2) the total number (or expected number) of required runs. These approximations compared very favorably with corresponding Monte Carlo estimates. More importantly, the approximations provide the user with quantitative information on the tradeoffs involved in particular screening applications. Accordingly, the results of this paper can be used as a practical guide in making objective decisions about the use of the two strategies we have studied.

Although other supersaturated screening strategies, such as group screening ([7], [11]), have been suggested, there has been little or no systematic evaluation of the performance of these methods. It remains to be seen, therefore, how the performance of these methods compare with the performance of the two strategies examined here.

APPENDIX

If H is a discrete random variable having probability distribution

$$P(H=h) = \begin{cases} \binom{r}{h}^2 \Big/ \binom{2r}{r} & h=0, 1, 2, \ldots, r \\ 0 & \text{otherwise,} \end{cases} \qquad (A.1)$$

where r is any positive integer, we write $H \sim H(r)$. The class of distributions defined in (A.1) is a symmetric subfamily of the hypergeometric family of distributions. If $H \sim H(r)$, then $E(H) = r/2$ and $V(H) = (r/2)^2/(2r-1)$.

We define $\underline{f}_i = (\underline{x}_i + \underline{1})/2$ where $\underline{x}_i$ is the $i^{\underline{th}}$ column vector of an RB design matrix and note that, for $i \neq j$, $\underline{f}_j'\underline{f}_i \sim H(N/2)$. It follows that, for $i \neq j$, $(\underline{x}_j'\underline{x}_i + N)/4 \sim H(N/2)$. Hence, for $i \neq j$, $E(\underline{x}_j'\underline{x}_i) = 0$ and $V(\underline{x}_j'\underline{x}_i) = N^2/(N-1)$.

In regard to the distribution of $\underline{x}_j'\underline{\varepsilon}$, we observe that $\underline{x}_j'\underline{\varepsilon}|\underline{x}_j \sim N(0, N\sigma^2)$ since $\underline{x}_j'\underline{x}_j = N$. Because the conditional distribution of $\underline{x}_j'\underline{\varepsilon}$ given $\underline{x}_j$ is the same for any realization of $\underline{x}_j$, the result is therefore true unconditionally. Thus, $E(\underline{x}_j'\underline{\varepsilon}) = 0$ and $V(\underline{x}_j'\underline{\varepsilon}) = N\sigma^2$.

To find the covariance between $\hat{\beta}_i$ and $\hat{\beta}_j$, for $i \neq j$, we have

$$\operatorname{cov}(\hat{\beta}_i, \hat{\beta}_j) = E(\hat{\beta}_i\hat{\beta}_j) - \beta_i\beta_j \ . \qquad (A.2)$$

We write $N\hat{\beta}_m = \gamma_m + \lambda_m$ where $\lambda_m = \underline{x}_m'\underline{\varepsilon}$ and $\gamma_m = \sum_{q=1}^{K} \beta_q \underline{x}_m'\underline{x}_q$. Thus,

$$N^2E(\hat{\beta}_i\hat{\beta}_j) = E(\gamma_i\gamma_j) + E(\gamma_i\lambda_j) + E(\gamma_j\lambda_i) + E(\lambda_i\lambda_j). \qquad (A.3)$$

It is easy to show that $E(\gamma_i\lambda_j) = E(\gamma_j\lambda_i) = E(\lambda_i\lambda_j) = 0$, so that

$$\operatorname{cov}(\hat{\beta}_i, \hat{\beta}_j) = E(\gamma_i\gamma_j)/N^2 - \beta_i\beta_j \ . \qquad (A.4)$$

Expanding $E(\gamma_i\gamma_j)$, we have

$$E(\gamma_i\gamma_j) = \sum_{r=1}^{K}\sum_{q=1}^{K} \beta_r\beta_q\, E(\underline{x}_i'\underline{x}_r\underline{x}_j'\underline{x}_q) \quad . \tag{A.5}$$

It is not difficult to verify that, for $i\neq j$, $E(\underline{x}_i'\underline{x}_r\underline{x}_j'\underline{x}_q)$ is zero unless $r=i$, $q=j$ or $r=j$, $q=i$. When $r=i$ and $q=j$, $E(\underline{x}_i'\underline{x}_r\underline{x}_j'\underline{x}_q) = E[(\underline{x}_i'\underline{x}_i)^2] = N^2$; when $r=j$ and $q=i$, $E(\underline{x}_i'\underline{x}_r\underline{x}_j'\underline{x}_q) = V(\underline{x}_j'\underline{x}_i) = N^2/(N-1)$.

Substitution into (A.5) yields

$$E(\gamma_i\gamma_j) = N^3\beta_i\beta_j/(N-1). \tag{A.6}$$

Finally, introducing (A.6) into (A.4), we get the desired result

$$\mathrm{cov}(\hat{\beta}_i,\hat{\beta}_j) = \beta_i\beta_j/(N-1) \quad . \tag{A.7}$$

REFERENCES

[1] Anscombe, F. J., "Quick Analysis Methods for Random Balance Screening Experiments," Technometrics, Vol. 1, pp. 195-209, 1959.

[2] Daniel, C., "Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments," Technometrics, Vol. 1, pp. 311-341, 1959.

[3] Feller, W., An Introduction to Probability Theory and Its Applications, Vol. 1, Third Edition, John Wiley & Sons, Inc., New York, 1968.

[4] Kleijnen, J. P. C., "Screening Designs for Poly-Factor Experimentation," Technometrics, Vol. 18, pp. 487-493, 1975.

[5] Mauro, C. A., and Smith, D. E., "Factor Screening in Simulation: Performance of a Two-Stage Random Balance/Plackett-Burman Procedure," Technical Report No. 113-5, Desmatics, Inc., 1981.

[6] Mauro, C. A., and Smith, D. E., "Factor Screening in Computer Simulation," Simulation, Vol. 38, pp. 49-54, 1982.

[7] Mauro, C. A., and Smith, D. E., "The Performance of Two-Stage Group Screening In Factor Screening Experiments," Technometrics, Vol. 24, 1982 (to appear).

[8] Montgomery, D. C.,"Methods for Factor Screening in Computer Simulation Experiments," Technical Report, Georgia Institute of Technology, 1979.

[9] Plackett, R. L., and Burman, J. P., "The Design of Optimum Multifactor Experiments," Biometrika, Vol. 33, pp. 305-325, 1946.

[10] Satterthwaite, F. E., "Random Balance Experimentation," Technometrics, Vol. 1, pp. 111-137, 1959.

[11] Watson, G. S., "A Study of the Group Screening Method," Technometrics, Vol. 3, pp. 371-388, 1961.

[12] Youden, W. J., Kempthorne, O., Tukey, J. W., Box, G. E. P., and Hunter, J. S., "Discussion of the Papers of Messrs. Satterthwaite and Budne," Technometrics, Vol. 1, pp. 157-193, 1959.

	E(R)	$j \leq 10$		$j \geq 11$	
		$P(R_{1j}=1)$	$P(R_{2j}=1)$	$P(R_{1j}=1)$	$P(R_{2j}=1)$
$\alpha_1 = .1$					
Approximation	43.25	.475	.452*,.451**	.100	.005*,.005**
Monte Carlo	42.86(.15)	.459(.006)	.432(.007)	.009(.001)	.005(.001)
$\alpha_1 = .3$					
Approximation	67.71	.721	.712*,.712**	.300	.015*,.015**
Monte Carlo	67.54(.23)	.720(.005)	.713(.006)	.299(.001)	.015(.001)
$\alpha_1 = .5$					
Approximation	90.89	.839	.836*,.836**	.500	.025*,.025**
Monte Carlo	90.61(.27)	.838(.005)	.837(.005)	.497(.001)	.024(.001)
$\alpha_1 = .7$					
Approximation	113.65	.915	.914*,.914**	.700	.035*,.035**
Monte Carlo	113.54(.23)	.913(.004)	.913(.004)	.699(.001)	.035(.001)

*Based on Bernoulli convolution approximation.
**Based on Poisson approximation.

Case Study I: K = 120 Factors; N = 24 Runs; $\alpha_1 = .1, .3, .5, .7$; $\alpha_2 = .05$

$$|\beta_j|/\sigma = \begin{cases} 4. & \text{for } j \leq 10 \\ 0. & \text{for } j \geq 11 \end{cases}$$

Table 1: Results of Case Study I. (Value in parentheses represents estimated standard deviation of Monte Carlo estimate.)

		$P(R_{1j}=1)$		$P(R_{2j}=1)$		E(R)	
		Monte Carlo	Approx.	Monte Carlo	Approx.*	Monte Carlo	Approx.
$\alpha_1 = .1$						48.00(.20)	48.33
$\lvert\beta_j\rvert/\sigma$	No.						
0.	30	.099(.003)	.100	.005(.001)	.005		
1.	24	.112(.004)	.109	.072(.003)	.071		
2.	18	.123(.004)	.136	.105(.004)	.117		
3.	12	.182(.006)	.180	.167(.006)	.165		
4.	9	.238(.008)	.242	.226(.007)	.230		
5.	5	.318(.011)	.320	.310(.011)	.311		
6.	2	.381(.020)	.412	.375(.020)	.406		
$\alpha_1 = .3$						69.43(.25)	69.82
$\lvert\beta_j\rvert/\sigma$	No.						
0.	30	.297(.005)	.300	.015(.001)	.015		
1.	24	.315(.005)	.313	.253(.006)	.251		
2.	18	.335(.006)	.351	.309(.006)	.322		
3.	12	.406(.008)	.410	.390(.008)	.395		
4.	9	.489(.009)	.486	.483(.009)	.480		
5.	5	.573(.013)	.573	.571(.013)	.571		
6.	2	.638(.020)	.664	.637(.020)	.664		

*Based on Poisson approximation.

Case Study II: K = 100 Factors $\alpha_1 = .1$ and .3

N = 30 Runs $\alpha_2 = .05$

Table 2: Results of Case Study II. (Value in parentheses represents estimated standard deviation of Monte Carlo estimate.)

Level of Significance	N=20 (c=.1)	N=40 (c=.2)	N=60 (c=.3)	N=80 (c=.4)	N=100 (c=.5)	N=120 (c=.6)	N=140 (c=.7)	N=160 (c=.8)	N=180 (c=.9)
$\alpha_1 = .05$	.088	.133	.178	.223	.268	.312	.356	.398	.438
$\alpha_1 = .10$	.157	.217	.275	.330	.383	.433	.480	.524	.565
$\alpha_1 = .15$	.218	.287	.351	.410	.465	.516	.563	.606	.645
$\alpha_1 = .20$	.276	.348	.415	.475	.530	.580	.625	.666	.703
$\alpha_1 = .25$	.329	.404	.471	.531	.585	.633	.675	.713	.747
$\alpha_1 = .30$	.381	.456	.521	.580	.631	.677	.717	.752	.783
$\alpha_1 = .35$	.431	.504	.567	.623	.672	.715	.752	.785	.813
$\alpha_1 = .40$	.479	.549	.610	.663	.709	.748	.783	.813	.838

Table 3: Powers For An Effect of Magnitude 3σ For Given $RB(N,\alpha_1)$ Strategy In Illustrative Situation.

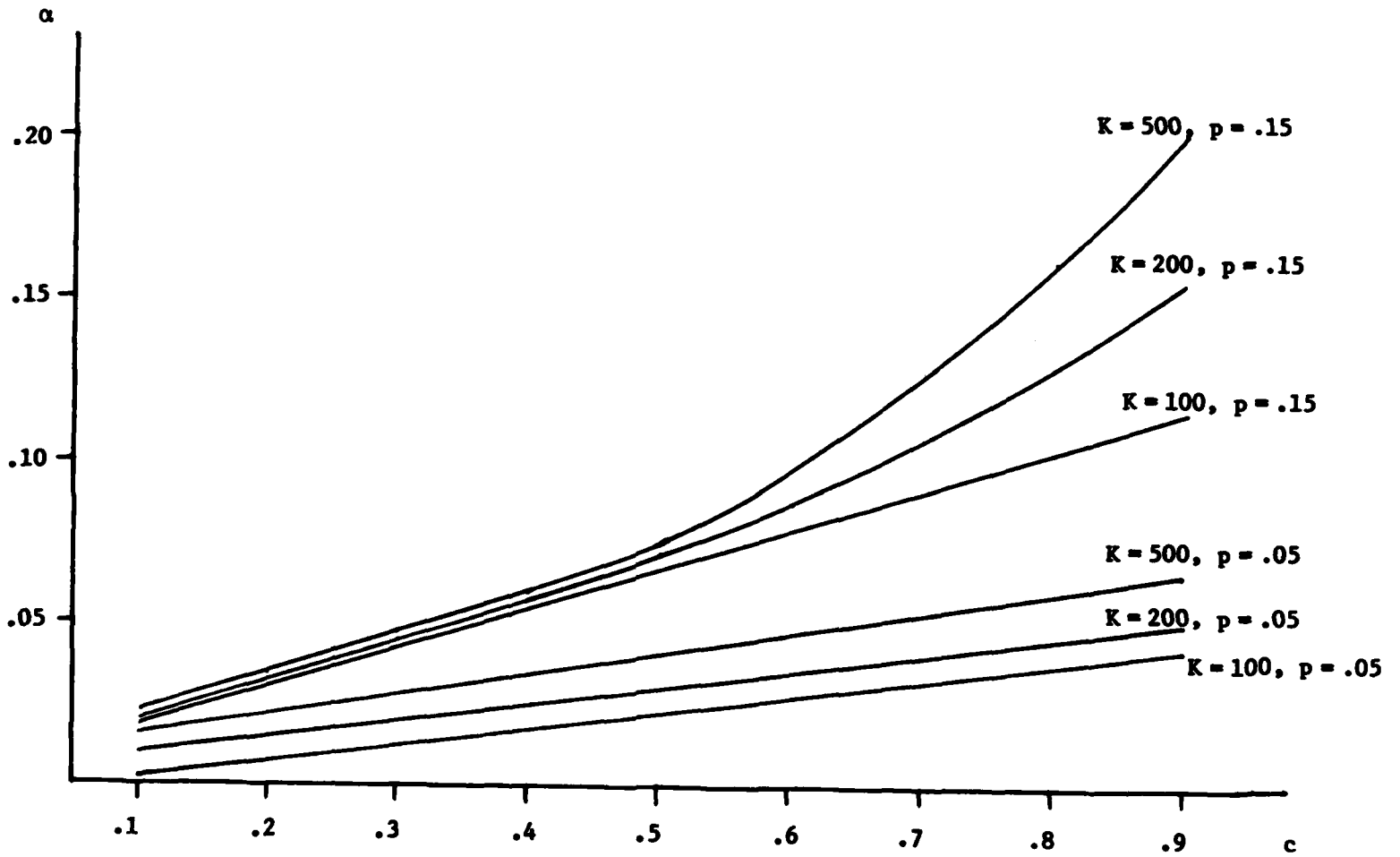


Figure 1: Curves Defining "Zones of Inadmissibility." Boundaries Are Accurate to ±.015 .

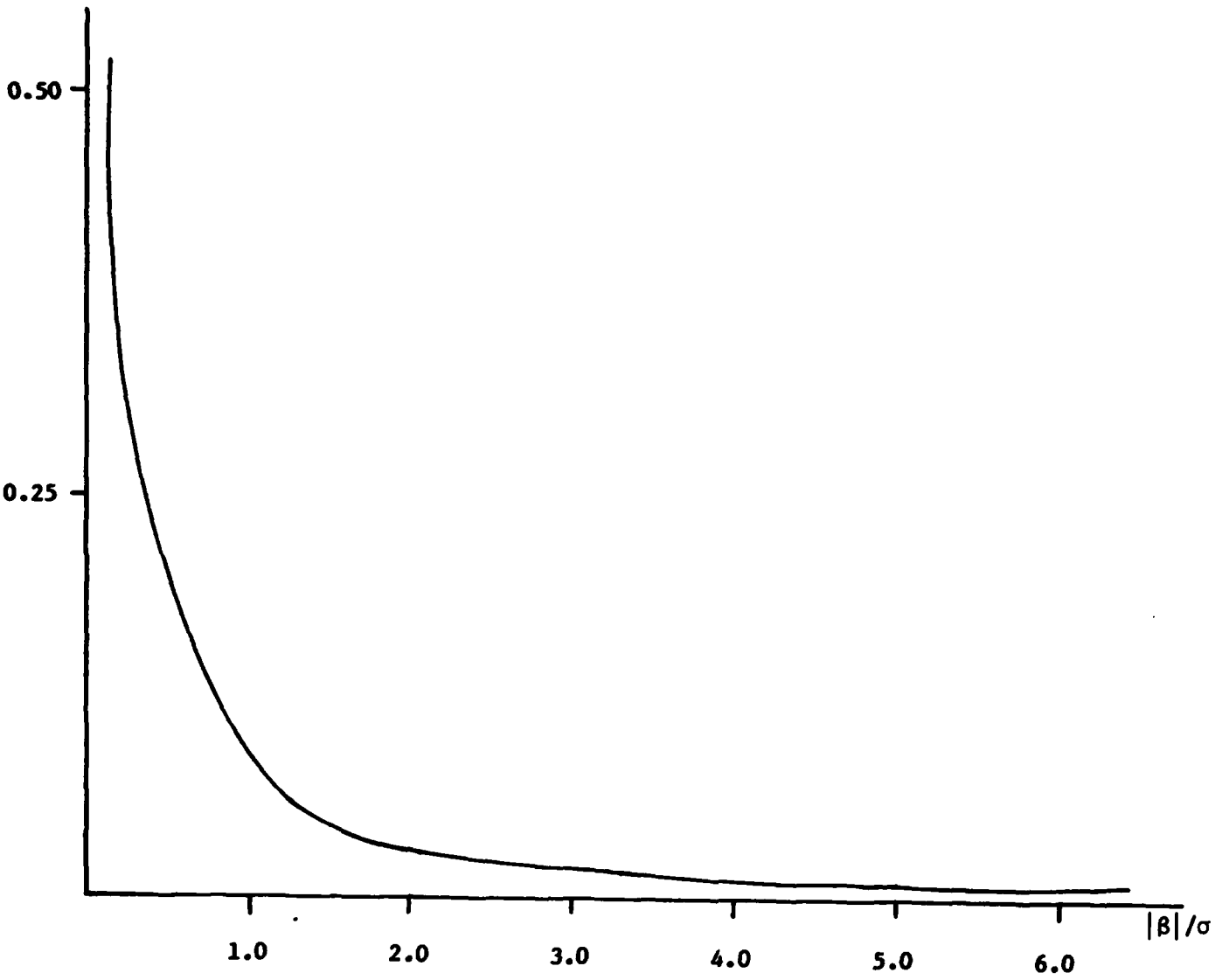


Figure 2: Hypothetical Distribution of Absolute Effects.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 113-7	2. GOVT ACCESSION NO.: AD-A121874
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): FACTOR SCREENING IN SIMULATION: EVALUATION OF TWO STRATEGIES BASED ON RANDOM BALANCE SAMPLING	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Carl A. Mauro and Dennis E. Smith	8. CONTRACT OR GRANT NUMBER(s): N00014-79-C-0650
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Desmatics, Inc. P.O. Box 618 State College, PA 16801	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 042-467
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research Arlington, VA 22217	12. REPORT DATE: November 1982
	13. NUMBER OF PAGES: 27
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this report is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Factor Screening
Computer Simulation
Simulation Methodology
Random Balance Sampling
Statistical Techniques in Simulation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In the study of large, complex computer simulation models the user is often overwhelmed by the vast number of input variables. Moreover, he or she is usually confused about how to make an effective analysis of the model without performing an excessive number of runs, which tend to be costly and time consuming. Factor screening methods, which attempt to identify the more important variables, can be extremely useful in the study of such models. This paper presents and evaluates two screening strategies based upon random balance sampling. Both strategies are applicable when there are more variables to be screened than

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

there are available screening runs. The results provide guidance in using these strategies in particular screening applications.

END

FILMED

1-83

DTIC